A Survey on Vision-based Fall Detection

Zhong Zhang, Christopher Conly, and Vassilis Athitsos
Department of Computer Science and Engineering
University of Texas at Arlington
Arlington, Texas, USA
zhong.zhang@mavs.uta.edu, cconly@uta.edu, athitsos@uta.edu

ABSTRACT

Falls are a major cause of fatal injury for the elderly population. To improve the quality of living for seniors, a wide range of monitoring systems with fall detection functionality have been proposed over recent years. This article is a survey of systems and algorithms which aim at automatically detecting cases where a human falls and may have been injured. Existing fall detection methods can be categorized as using sensors, or being exclusively vision-based. This literature review focuses on vision-based methods.

Categories and Subject Descriptors

J.3 [**LIFE AND MEDICAL SCIENCES**]: Health

General Terms

Survey

Keywords

Fall detection, Kinect, Survey

1. INTRODUCTION

Falls are a major public health problem for the elderly population because they cause many injuries or even death. To avoid these tragedies that arise from falls, quick responses (alerting relatives and/or authorities) are important when falls happen. It has been proved that the earlier a fall is reported, the higher probability the patient would recover from the accident. At the same time, the fear of falling also prevents elderly people and patients from living unaccompanied at home which absolutely increases manual labor in terms of the presence of nurses or support staff. Based on the above facts, the demand for developing some intelligent monitoring systems with the ability to detect falls automatically has boomed in the healthcare industry. An intelligent surveillance system, which is capable of detecting falls accurately, can not only improve the quality of living for the elderly, but also save on manual labor. The level of reliability and practicality of a surveillance system heavily depends on its fall detection module. Research has already been done to design algorithms to detect falls for many years. Existing approaches can be broadly divided into two groups: using a variety of non-vision sensors (the most commonly used sensors are accelerometers), and being exclusively vision-based. Generally speaking, the sensor-based methods require subjects to actively cooperate by wearing the sensors, which can be problematic and possibly uncomfortable (e.g., wearing sensors while sleeping, to detect falls during a night trip to the restroom). The vision-based methods are less intrusive, as all information is collected from cameras.

PETRA '15, July 01 - 03 2015, Island of Corfu, Greece.
 ISBN 978-1-4503-3452-5/15/07…$15.00.
DOI: http://dx.doi.org/10.1145/2769493.2769540

This literature review focuses on vision-based research work and contains a comprehensive study of recent proposed fall detection methods using depth cameras. Compared to existing review papers [27, 18], this literature review has the following contributions: (1) we focus on recent vision-based fall detection techniques. Specifically, the recent depth cameras based fall detection methods are extensively summarized in this survey. (2) we are not aware of any literature discussing the publicly available fall datasets. However, establishing several benchmark datasets is extremely important for the fall detection community, which enables researchers to fairly compare their methods with others. This literature review introduces several publicly available fall datasets that can serve as benchmarks.

The rest of the paper is organized as follows. In Section 2, several publicly available fall datasets are introduced. We first talk about the classification of vision-based fall detection methods and then introduce them separately in Section 3. Finally, we conclude and discuss future directions of research in Section 4.

2. PUBLIC FALL DATASETS

We introduce five publicly available fall datasets in this section. Three of them were recorded using Kinect cameras, one was collected by a single RGB camera and one was made with multiple calibrated RGB cameras.

SDUFall[1] [23]: A Kinect camera was set up to record the dataset. Three channels were collected: RGB video (.avi), depth video (.avi), and 20 skeleton joint positions (.txt). All videos were recorded at a resolution of 320x240 pixels per frame and 30 frames per second in AVI format. Twenty subjects participated in the data recording. Each subject performed 6 actions 10 times each: falling down, bend-

[1] `http://www.sucro.org/homepage/wanghaibo/SDUFall.html`

Table 1: Five publicly available fall datasets

	SDUFall	EDF	OCCU	Dataset introduced in [9]	Dataset introduced in [7]
camera type	one Kinect	two Kinects	two Kinects	one RGB camera	eight calibrated RGB cameras
camera viewpoints	one	two	two	NaN	eight
fall type	falls with different directions	eight fall directions	occluded falls	falls with different directions	forward, backward falls, falls from sitting down and loss of balance
number of falls	200	320	60	192	200
activities of daily life	Yes	Yes	Yes	Yes	Yes
simulated scenarios	1	1	1	4 (home, coffee room, office, lecture room)	24

ing, squatting, sitting, lying and walking. Each action was recorded under certain conditions. These conditions include carrying or not carrying large object, turning the light on or off, changing direction and position relative to the camera.

EDF[2]: Two Kinect cameras were set up to record the **EDF** dataset. The two viewpoints were recorded at the same time, and thus every event was recorded simultaneously from both viewpoints. There are eight fall directions in the **EDF** dataset. Each of the 10 subjects performed two falls along each direction in each viewpoint in the EDF dataset. So, there are 160 falls in each viewpoint and 320 falls in total. In the **EDF** dataset, subjects also performed a total of 100 actions that tend to produce features similar to those of a fall event, namely: 20 examples of picking up something from the floor, 20 cases of sitting on the floor and 20 examples of lying down on the floor, 20 examples of tying shoelaces and 20 examples of doing plank exercise. The dataset was recorded at a resolution of 320x240 pixels per frame and at a frame rate of about 25 frames per second.

OCCU[3] [41]: Two Kinect depth cameras were set up at two corners of a simulated apartment to collect occluded falls. An occluded fall refers to the end of the fall can be completely occluded by a certain object, like a bed. Each of the 5 subjects performed six occluded falls in each viewpoint in the **OCCU** dataset. The **OCCU** dataset includes 25,618 frames and 30 totally occluded falls in videos from the first viewpoint, and 23,703 frames and 30 totally occluded falls in videos from the second viewpoint performed by the same subjects. Each viewpoint was recorded at separate times from the other viewpoint, and thus we had no instances where the same events were recorded simultaneously from both viewpoints. The subjects also performed a total of 80 actions that tended to produce features similar to those of a fall event, namely: 20 examples of picking up something from the floor (all of them are non-occluded), 20 examples of sitting on the floor (all of them are non-occluded), 20 examples of tying shoelaces (all of them are non-occluded), and 21 examples of lying down on the floor (all of them are totally occluded at the end frame). Figure 2 shows an example of an occluded fall in each viewpoint.

In the work [9] a fall dataset was collected using a single RGB camera. The frame rate is 25 frames/s and the resolution is 320x240 pixels. The video sequences contain variable

[2]https://sites.google.com/site/kinectfalldetection/
[3]http://sites.google.com/site/occlusiondataset

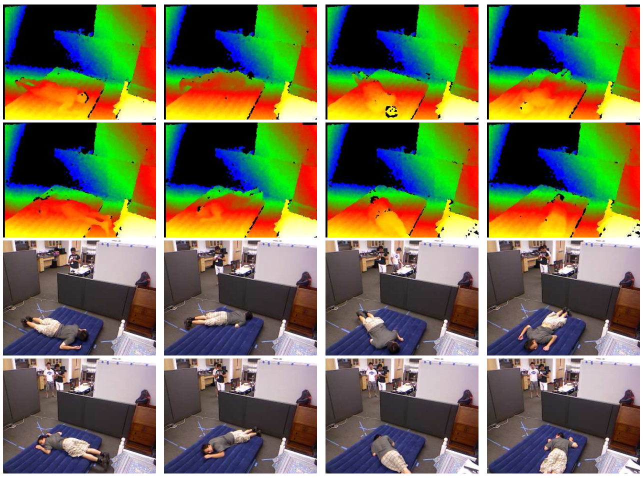

Figure 1: A person falls down along eight different directions in EDF dataset. The top two rows are shown as color images while the bottom two rows are shown as depth images. Depth images are color-coded so that: white indicates small depth values, and yellow, orange, red, green, blue indicate progressively larger depth values. Black indicates invalid depth.

illumination, and typical difficulties like occlusions or cluttered and textured background. The actors performed various normal daily activities and falls. The dataset contains 250 videos and associated annotations marking beginning and end of each fall event and indicating the location of the human body in each frame. In order to evaluate the robustness of the method to the location change between traning and testing, the dataset was recorded from different locations (“Home”, “Coffee room”, “Office” and “Lecture room”).

Eight inexpensive IP cameras with a wide angle were set up to cover the whole room in [7]. The collected dataset is composed of several simulated normal daily activities and falls viewed from all the cameras and performed by one subject. Normal daily activities include walking in different directions, housekeeping, activities with characteristics similar to falls (sitting down/standing up, crouching down). Simulated falls include forward falls, backward falls, falls when inappropriately sitting down, loss of balance. Falls were performed in different directions with respect to the camera point of view.

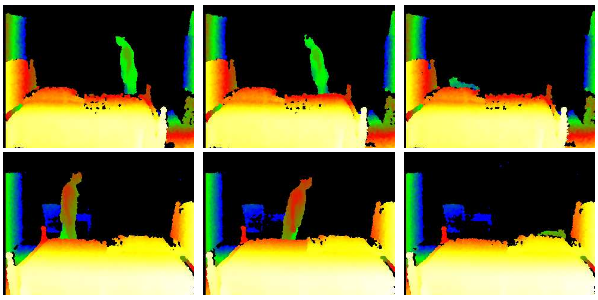

Figure 2: Examples of the occluded fall. The top row shows an occluded fall in the first viewpoint while the bottom row shows an occluded fall in the second viewpoint.

3. VISION-BASED FALL DETECTORS

Cameras are increasingly included in home assistive/care systems as they possess multiple advantages over other sensor-based systems. Cameras can be used to detect multiple actions simultaneously with less intrusion. Vision-based methods can be broadly divided into three categories: fall detection using a single RGB camera, 3D-based methods using multiple cameras, and 3D-based methods using depth cameras.

3.1 Fall Detection Using a Single RGB Camera

Fall detections using a single RGB camera have been extensively studied, as the systems are easy to set up and are inexpensive. Shape related features, inactivity detection and human motion analysis are the most commonly used clues for detecting falls.

Shape related features are widely used for fall detection [26, 34, 38, 25, 36, 14, 29]. In [36, 25], fall detection is based on width to height aspect ratio of the person. Mirmahboub et al. [26] use a simple background separation method to create the silhouette of the person, and several features are then extracted from the silhouette area. Finally, A SVM classifier is employed to perform the classification based on these silhouette-related features. Rougier et al. [34] use a shape matching technique to track the silhouette of the person in the target video clip. The shape deformation is then quantified from these silhouettes, and the classification is based on the shape deformation using a Gaussian mixture model. An adaptive background Gaussian mixture model (GMM) is employed to obtain the moving object in [38], and an ellipse shape is built from the moving object for body modeling. Several features are then extracted from the ellipse model. Unlike [34], two Hidden Markov Models (HMMs) are used to classify falls and normal activities. Arie et al. [29] present a novel method to distinguish different postures including standing, sitting, bending/squatting, lying on the side and lying toward the camera. The proposed method extracts the projection histograms of the segmented human body silhouette as the main feature vector. Posture classification is completed by k-Nearest Neighbor (k-NN) algorithm and evidence accumulation technique.

The motion pattern differences between falls and other daily activities, like walking, sitting down, drinking and etc, are significant. Much of the research is based on motion analysis [15, 22, 40, 33]. Liao et al. [22] use human motion analysis and human silhouette shape variations to detect slip-only and fall events. The motion measure is obtained by analyzing the energy of the motion active (MA) area in the integrated spatio-temporal energy (ISTE) map. Homa et al. [15] discuss applying Integrated Time Motion Image (ITMI) to fall detection. Integrated Time Motion Image (ITMI) is a type of spatio-temporal database that includes motion and time of motion occurrence. Given a video clip, the integrated time motion images are calculated to represent the motion pattern that occurred in the video and then PCA is used for feature reduction. Finally, a pre-trained MLP Neural Network is adopted for precise classification of motions and determination of a fall event. Zhang et al. [40] describe experiments with three computer vision methods for fall detection in a simulated home environment. The first method makes a decision based on a single frame, simply based on the vertical position of the image centroid of the person. The second method makes a threshold-based decision based on the last few frames, by considering the number of frames during which the person has been falling, the magnitude (in pixels) of the fall, and the maximum velocity of the fall. The third method is a statistical method that makes a decision based on the same features as the second method, but using probabilistic models as opposed to thresholds for making the decision. Caroline et al. [33] extract the 3D head trajectory using a single calibrated camera. With the help of 3D head trajectory, the velocity characteristics are calculated for fall detection. Feng et al. [14] propose a novel vision-based fall detection method for monitoring elderly people in a house care environment. The foreground human silhouette is extracted via background modeling and tracked throughout the video sequence. The human body is represented with ellipse fitting, and the silhouette motion is modeled by an integrated normalized motion energy image computed over a short-term video sequence. Then, the shape deformation quantified from the fitted silhouettes is used as the features to distinguish different postures of the human.

Inactivity detection is adopted by [28] to detect falls. In [28] ceiling-mounted, wide-angle cameras with vertically oriented optical axes are used to reduce the influence of occlusion. Nait-Charif et al. [28] use learned models of spatial context, which are used in conjunction with a tracker to achieve these goals. Nater et al. [30] present an approach for unusual event detection based on a tree of trackers. Each tracker is specialized for a specific type of activity. Falls are detected when none of the specialized trackers for "normal" activities can explain the observation. Charfi et al. [11, 10] introduce a spatio-temporal human fall descriptor, named STHF, that uses several combinations of transformations of geometrical features. The well-known SVM classifier is applied to the STHF descriptor to classify falls and normal activities.

3.2 3D-based Methods Using Multiple RGB Cameras

Another category of vision-based methods for fall detection is 3D-based methods using multiple RGB cameras. The calibrated multi-camera systems [6, 4, 5, 1, 2] allow 3D reconstruction of the object but require a careful and time-consuming calibration process. Auvinet et al. [6, 4] use a network of multiple calibrated cameras to reconstruct the 3D shape of the person. Fall events are detected by analyzing the volume distribution along the vertical axis, and an alarm is triggered when the major part of this distribution is abnormally near the floor. In a later work [5], the fall alarm would be triggered when the major part of this distribution is abnormally near the floor during a predefined period of time. Anderson et al. [1, 2] employ multiple cameras and a hierarchy of fuzzy logic to detect falls. Overall, using multiple cameras offers the advantage of allowing 3D reconstruction and extraction of 3D features for fall detection. The proposed method introduces the voxel person, which is a linguistic summarization of temporal fuzzy inference curves, to represent the states of a three-dimensional object.

Aside from providing 3D reconstruction, multi-camera systems can also be used for other purposes, like viewpoint independent [37] fall detection, monitoring multiple rooms [12] and fusion of different cameras' results [44]. Thome et al. [37] propose a multi-view fall detection system by which motion is modeled by a layered hidden Markov model (LHMM). The proposed method uses a multi-view setting, where the low-level steps are (mainly) performed independently in each view, leading to the extraction of simple image features compatible with real-time achievement. Then, a fusion unit merges the output of each camera to provide a viewpoint-independent pose classification. Cucchiara et al. [12] use multiple cameras to monitor different rooms. A single room is monitored by a single camera. Multiple cameras are used to cover different rooms and the camera handoff is treated by warping the person's appearance in the new view by means of homography. Zweng et al. [44] detect falls using multiple cameras. Each of the camera inputs results in a separate fall confidence (so no external camera calibration is needed). These confidences are then combined into an overall decision. Hung et al. [16, 17] propose using the measures of humans' heights and occupied areas to distinguish three typical states of humans: standing, sitting and lying. Two relatively orthogonal views are utilized, in turn, simplifying the estimation of occupied areas as the product of widths of the same person, observed in two cameras.

3.3 3D-based Methods Using Depth Cameras

The earliest depth camera used for fall detection is the Time-Of-Flight 3D camera [13]. Since the price of a Time-Of-Flight 3D camera is expensive, very few researchers adopted it for fall detection. But this situation has changed since the advent of the affordable depth sensing technology, like Microsoft Kinect.

With the help of the depth cameras, the calculation of the distance from the top of the person to the floor is simple, which can then be used as a feature to detect falls [13, 21, 32, 19]. Diraco et al. [13] use a wall-mounted Time-Of-Flight 3D camera to monitor the scene. The system identifies a fall event when the human centroid gets closer than a certain threshold to the floor, and the person does not move for a certain number of seconds once close to the floor. In a related approach, Leone et al. [21] employ a 3D range camera. A fall event is detected based on two rules: (1) the distance of the person's center-of-mass from the floor plane decreases below a threshold within a time window of about $900ms$; (2) the person's motion remains negligible within a time window of about $4s$. Rougier et al. [32] use a Kinect camera to obtain depth images. Thresholds on the human centroid height relative to the ground and the body velocity are used to determine if a fall has occurred. Michal et al. [19] use a ceiling-mounted 3D depth camera to detect falls. A KNN classifier is used to distinguish the lying pose from common daily activities based on features including head to floor distance, person area and shape's major length to width. Human motion analysis is further employed to classify between intentional lying postures and accidental falls.

Analyzing how a human has moved during the last frames in a world coordinate system is another commonly used method [35, 42, 41, 24]. Eric et al. [35] propose a two-stage fall detection method. The first stage of the system characterizes the vertical state of a segmented 3D object for each frame, and then identifies on ground events through temporal segmentation of the vertical state time series of tracked 3D objects. The second stage of the system utilizes an ensemble of decision trees and features extracted from an on ground event to compute a confidence that a fall preceded it. Zhong et al. [42] propose a statistical method based on Kinect depth cameras, that makes a decision based on information about how the human moved during the last few frames. The proposed method introduces five novelty features to be used for fall detection, including duration, total head drop, maximum speed, smallest head height and fraction of frames for which the head drops, and combines these features using a Bayesian framework. The proposed method adopts a viewpoint independent experimental protocol, whereby all training data are collected from a specific viewpoint, and all test data are collected from another viewpoint, several meters away from the training viewpoint. This evaluation protocol is a useful approach for measuring the robustness of the system to displacements of the camera. In a later work, Zhong et al. [41] discuss the topic of occlusion falls. They define a complete occlusion fall as one where the end of the fall is totally occluded by a certain object, like a bed. To quantify the challenges and assess performance in this topic, the authors present an occluded fall detection benchmark dataset containing 60 occluded falls for which the end of the fall is completely occluded. They also evaluate four existing fall detection methods using a single depth camera on this benchmark dataset. Georgios et al. [24] present a novel fall detection system based on the Kinect sensor. The fall is initialized by the analysis of the velocity change and then finalized by the inactivity detection. The key novelty of the proposed method lies in calculating the velocity based on the contraction or expansion of the width, height and depth of the 3D bounding box. By explicitly using the 3D bounding box, the proposed algorithm does not require any prior knowledge of the scene (i.e. floor). The inactivity situation is defined as a lack of motion for the monitored human in a pre-defined time window.

The Microsoft Kinect SDK provides skeletal joints tracking, which enables researchers to analyze the human body key joints for fall detection [8, 39]. Zhen-Peng et al. [8] propose a single depth camera based fall detection method by analyzing the human body key joints. In the proposed approach, a pose-invariant and efficient randomized decision tree (RDT) algorithm is employed to extract the 3D body joints at each frame. Then, the 3D trajectory of the head joint is fed to a pre-trained support vector machine (SVM) classifier to determine whether or not a fall action has occured. Chenyang et al. [39] propose a RGB-D cameras based method to recognize five activities: standing, falling from standing, falling from chair, sitting on chair, and sitting on floor. The main analysis is based on 3D depth information. If the person goes out of the range of the depth camera, RGB video is used to analyze the activities. The kinematic model features that are extracted from 3D depth information include two components: structure similarity and vertical height of the person. For RGB video analysis, the width-height ratio of the detected human bounding box is used to recognize different activities.

Other methods using depth cameras include [23, 3, 20]. Xin et al. [23] combine shape-based fall characterization and a learning-based classifier to distinguish falls from other daily actions. Curvature scale space (CSS) features of human silhouettes are extracted at each frame and then an action is represented by a bag of CSS words (BoCSS). The BoCSS representation of a fall is distinguished from those of other actions by the pre-trained extreme learning machine (ELM) classifier. In a later work [3], instead of representing an action as a bag of CSS words, Fisher Vector (FV) encoding is used to describe the action based on CSS features. A pre-trained Support Vector Machine (SVM) classifier is employed to do the final classification. Lee et al. [20] present a rules based abnormal event detection method using both color and depth image. The proposed algorithm exploits three features, including the width-height ratio of the detected human bounding box, normalized 2D velocity, and 3D depth information. More specifically, if the width-height ratio of the bounding box is greater than a pre-defined threshold, the system reports a fall event. Those cases, where the width-height ratio of the bounding box is less than the predefined threshold, are further checked by analyzing the 2D velocity, which is calculated as the motion of the centroid of the human bounding box, and the 3D centroid information.

4. CONCLUSION AND FUTURE WORK

We have studied different vision-based methods for detecting fall events. One attractive feature of single RGB camera based methods is that they do not require camera calibration. Furthermore, the setup is easy, and they are inexpensive. The disadvantage is that most of them lack flexibility. Fall detectors using a single RGB camera are often case specific and viewpoint-dependent. Moving a camera to a different viewpoint (especially a different height from the floor) would require the collection of new training data for that specific viewpoint. Single RGB camera based methods also suffer from occlusion problems where part of the human body is occluded by a certain object, like a bed.

Calibrated multi-camera systems enable researchers to reconstruct 3D information of the person and are, thus, inherently viewpoint independent. Setting multiple cameras around the environment to make sure there is no hidden area would be helpful for solving the occlusion problems that happen in single camera scenario. However, calibrated multi-camera systems require a careful and time-consuming calibration process and we should have in mind that the calibration process needs to be repeated every time a single camera is intentionally or accidentally moved.

Depth camera based systems share the same advantages with the calibrated multi-camera systems and at the same time do not need a time-consuming calibration step. Depth cameras enable researchers to model fall events under a camera-independent world coordinate system and reconstruct the 3D information of the object. In addition, Kinect, which is the most commonly used depth camera, can be used to track human skeletal joints. Several fall detection methods are based on analyzing the change of the skeletal joints. Although the cost of a depth camera is more expensive than a RGB camera, the price is still affordable.

From a research perspective, it is important to have a publicly available benchmark dataset of falls which enables researchers to evaluate their fall detectors and fairly compare their methods with other detectors. A comprehensive fall dataset should include the following properties:

- The main objective of a fall detector is to discriminate between fall events and activities of daily living (ADL). This is not a trivial task as certain ADL, liking sitting down or going from standing position to lying down, have similar features to falls. Thus, in order to test the quality of a fall detector, it is necessary to collect both falls and ADL.

- Different scenarios should be considered when collecting the fall dataset. There are different kinds of falls: forward or backward falls, falls from sleeping in the bed or sitting on a chair, falls from standing on supports or walking, and etc. These different falls can exhibit significantly different properties while also sharing some common characteristics. To include various falls in the benchmark dataset would make it comprehensive and practical, as well as present challenges to researchers and thus push them to design better algorithms.

- It is also helpful to include different camera viewpoints in the benchmark dataset. A practical fall detection system should be viewpoint independent which means the fall detection algorithm does not depend on any specific camera viewpoint and moving the camera should not affect the system. Different camera viewpoints in the dataset can be used to verify whether the proposed algorithms have this viewpoint-independent property.

- It is very difficult to have real falls because of the privacy issue. So, the falls are generally simulated by volunteers which is a feasible option adopted by most researchers. When recording simulated falls, we should consider the distribution of the volunteers: male and female, different ages, etc.

We are not aware of any existing publicly available fall datasets (Section 2) satisfying all the above conditions. So, making a comprehensive fall dataset that has all the above properties is urgent and would be of benefit to the research community.

A comprehensive and robust fall detection system should be both high sensitivity and good specificity. Unfortunately, the existing vision-based fall detection methods cannot satisfy both accuracy and robustness. Although we can develop more sophisticated vision-based techniques, it is also worth emphasizing that the vison-based methods are not necessary to be a standalone system for fall detection, can also be combined with other modules to form a general system. The overall fall detection system may contain additional modules, both to improve accuracy, and to include additional functionality, such as an acoustic module [31, 43] and a module sending an alert about the detected fall. We believe that including sound processing and speech recognition would help significantly towards obtaining a robust system. Sound processing may produce additional features to be used for classifying a candidate fall event. Speech recognition can be used so that the system initiates a dialog with the subject, in the case where a fall has been detected. For example, the system can ask “Are you OK?” and the user can respond to indicate that there was no actual fall and no need to issue an alert.

5. ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grants IIS-1055062, CNS-1059235, CNS-1035913, and CNS-1338118.

6. REFERENCES

[1] D. Anderson, R. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer Vision and Image Understanding*, 113(1):80–89, Jan 2009.

[2] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. J. Rantz, and M. A. Aud. Modeling human activity from voxel person using fuzzy logic. *Fuzzy Systems, IEEE Transactions on*, 17(1):39–49, 2009.

[3] M. Aslan, A. Sengur, Y. Xiao, H. Wang, M. C. Ince, and X. Ma. Shape feature encoding via fisher vector for efficient fall detection in depth-videos. *Applied Soft Computing*, 2015.

[4] E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau, and J. Meunier. Fall detection using body volume reconstruction and vertical repartition analysis. In *International conference on Image and signal processing*, pages 376–383, 2010.

[5] E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau, and J. Meunier. Fall detection with multiple cameras: An occlusion-resistant method based on 3-d silhouette vertical distribution. *IEEE Transactions on Information Technology in Biomedicine*, 15:290–300, March 2011.

[6] E. Auvinet, L. Reveret, A. St-Arnaud, J. Rousseau, and J. Meunier. Fall detection using multiple cameras. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 2554–2557. IEEE, 2008.

[7] E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Multiple cameras fall dataset. *DIRO-Université de Montréal, Tech. Rep*, 1350, 2010.

[8] Z. Bian, J. Hou, L. Chau, and N. Magnenat-Thalmann. Fall detection based on body part tracking using a depth camera. *IEEE journal of biomedical and health informatics*, 2014.

[9] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki. Definition and performance evaluation of a robust svm based fall detection solution. In *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on*, pages 218–224. IEEE, 2012.

[10] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki. Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification. *Journal of Electronic Imaging*, 22(4):041106–041106, 2013.

[11] I. Charfi, J. Miteran, J. Dubois, B. Heyrman, and M. Atri. Robust spatio-temporal descriptors for real-time svm-based fall detection. In *Computer Applications & Research (WSCAR), 2014 World Symposium on*, pages 1–6. IEEE, 2014.

[12] R. Cucchiara, A. Prati, and R. Vezzani. A multi-camera vision system for fall detection and alarm generation. *Expert Systems*, 24(5):334–345, Nov 2007.

[13] G. Diraco, A. Leone, and P. Siciliano. An active vision system for fall detection and posture recognition in elderly healthcare. In *Design, Automation & Test in Europe Conference & Exhibition*, pages 1536–1541, March 2010.

[14] W. Feng, R. Liu, and M. Zhu. Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera. *Signal, Image and Video Processing*, 8(6):1129–1138, 2014.

[15] H. Foroughi, A. Naseri, A. Saberi, and H. S. Yazdi. An eigenspace-based approach for human fall detection using integrated time motion image and neural network. In *Signal Processing, 2008. ICSP 2008. 9th International Conference on*, pages 1499–1503. IEEE, 2008.

[16] D. Hung and H. Saito. The estimation of heights and occupied areas of humans from two orthogonal views for fall detection. *IEEJ Transactions on Electronics and Information and Systems*, 133, January 2013.

[17] D. H. Hung, H. Saito, and G.-S. Hsu. Detecting fall incidents of the elderly based on human-ground contact areas. In *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*, pages 516–521. IEEE, 2013.

[18] R. Igual, C. Medrano, and I. Plaza. Challenges, issues and trends in fall detection systems. *Biomed. Eng. Online*, 12(66):1–66, 2013.

[19] M. Kepski and B. Kwolek. Fall detection using ceiling-mounted 3d depth camera. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 640–647, 2014.

[20] Y.-S. Lee and W.-Y. Chung. Visual sensor based abnormal event detection with moving shadow removal in home healthcare applications. *Sensors*, 12(1):573–584, 2012.

[21] A. Leone, G. Diraco, and P. Siciliano. Detecting falls with 3d range camera in ambient assisted living applications: A preliminary study. *Medical Engineering & Physics*, 33:770–781, July 2011.

[22] Y. T. Liao, C.-L. Huang, and S.-C. Hsu. Slip and fall event detection using bayesian belief network. *Pattern*

recognition, 45(1):24–32, 2012.

[23] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li. Depth-based human fall detection via shape features and improved extreme learning machine. *IEEE journal of biomedical and health informatics*, 18(6):1915–1922, 2014.

[24] G. Mastorakis and D. Makris. Fall detection system using kinectâĂŹs infrared sensor. *Journal of Real-Time Image Processing*, 9(4):635–646, 2014.

[25] S. Miaou, P. Sung, and C. Huang. A customized human fall detection system using omni-camera images and personal information. In *Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare*, pages 39–42, Apr 2006.

[26] B. Mirmahboub, S. Samavi, N. Karimi, and S. Shirani. Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE Transactions on Biomedical Engineering*, 60:427–436, February 2013.

[27] M. Mubashir, L. Shao, and L. Seed. A survey on fall detection: Principles and approaches. *Neurocomputing*, 100:144–152, 2013.

[28] H. Nait-Charif and S. McKenna. Activity summarisation and fall detection in a supportive home environment. In *International Conference on Pattern Recognition*, pages 323–326, Aug 2004.

[29] A. H. Nasution and S. Emmanuel. Intelligent video surveillance for monitoring elderly in home environments. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*, pages 203–206. IEEE, 2007.

[30] F. Nater, H. Grabner, T. Jaeggli, and L. Gool. Tracker trees for unusual event detection. In *International Conference on Computer Vision Workshops*, pages 1113–1120, Sept 2009.

[31] M. Popescu, Y. Li, M. Skubic, and M. Rantz. An acoustic fall detector system that uses sound height information to reduce the false alarm rate. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 4628–4631. IEEE, 2008.

[32] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, and J. Meunier. Fall detection from depth map video sequences. In *International Conference on Smart Homes and Health Telematics*, pages 121–128, June 2011.

[33] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3d head tracking to detect falls of elderly people. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 6384–6387. IEEE, 2006.

[34] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Robust video surveillance for fall detection based on human shape deformation. *IEEE Transactions on Circuits and Systems for Video Technology*, 21:611–622, May 2011.

[35] E. Stone and M. Skubic. Fall detection in homes of older adults using the microsoft kinect. *IEEE journal of biomedical and health informatics*, 19:290–301, 2014.

[36] J. Tao, M. Turjo, M. Wong, M. Wang, and Y. Tan. Fall incidents detection for intelligent video surveillance. In *IEEE Conference on Information, Communications and Signal Processing*, pages 1590–1594, 2005.

[37] N. Thome, S. Miguet, and S. Ambellouis. A real-time, multiview fall detection system: A lhmm-based approach. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1522–1532, 2008.

[38] K. Tra and T. Pham. Human fall detection based on adaptive background mixture model and hmm. In *International Conference on Advanced Technologies for Communications*, pages 95–100, Oct 2013.

[39] C. Zhang, Y. Tian, and E. Capezuti. Privacy preserving automatic fall detection for elderly using rgbd cameras. In *Proceedings of the 13th International Conference on Computers Helping People with Special Needs - Volume Part I*, pages 625–633, 2012.

[40] Z. Zhang, E. Becker, R. Arora, and V. Athitsos. Experiments with computer vision methods for fall detection. In *International Conference on Pervasive Technologies Related to Assistive Environments*, pages 25:1–25:4, 2010.

[41] Z. Zhang, C. Conly, and V. Athitsos. Evaluating depth-based computer vision methods for fall detection under occlusions. In *Advances in Visual Computing*, pages 196–207. Springer, 2014.

[42] Z. Zhang, W. Liu, V. Metsis, and V. Athitsos. A viewpoint-independent statistical method for fall detection. In *IEEE Conference on Pattern Recognition*, pages 3626–3630, Nov 2012.

[43] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson. Acoustic fall detection using gaussian mixture models and gmm supervectors. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 69–72. IEEE, 2009.

[44] A. Zweng, S. Zambanini, and M. Kampel. Introducing a statistical behavior model into camera-based fall detection. In *International conference on Advances in visual computing*, pages 163–172, 2010.